A Structural Correlation Filter Combined with A Multi-task Gaussian Particle Filter for Visual Tracking

Manna Dai
Manna.Dai@student.uts.edu.au

Shuying Cheng
sycheng@fzu.edu.cn

Xiangjian He
Xiangjian.He@uts.edu.au

Dadong Wang
Dadong.Wang@data61.csiro.au

Abstract

In this paper, we propose a novel structural correlation filter combined with a multi-task Gaussian particle filter (KCF-GPF) model for robust visual tracking. We first present an hieraichical structure where several KCF trackers as weak experts provide a preliminary decision for a Gaussian particle filter to make a final decision. The proposed method is designed to exploit and complement the strength of a KCF and a Gaussian particle filter. Compared with the existing tracking methods based on correlation filters or particle filters, the proposed tracker has several advantages. First, it can detect the tracked target in a large-scale search scope via weak KCF trackers and evaluate the reliability of weak trackers'decisions for a Gaussian particle filter to make a strong decision, and hence it can tackle fast motions, appearance variations, occlusions and re-detections. Second, it can effectively handle large-scale variations via a Gaussian particle filter. Third, it can be amenable to fully parallel implementation using importance sampling without resampling, thereby it is convenient for VLSI implementation and can lower the computational costs. Extensive experiments on the OTB-2013 dataset containing 50 challenging sequences demonstrate that the proposed algorithm performs favourably against 16 state-of-the-art trackers.

1. Introduction

Visual tracking is one of the most fundamental problems in computer vision due to its numerous applications such as video surveillance, motion analysis, vehicle navigation and human computer interactions. Although a great progress has been seen on developing algorithms [19, 39, 40, 46] and benchmark evaluations [43] for visual tracking, visual tracking is still a challenging problem in the situations of heavy illumination changes, pose deformations, partial and full occlusions, large scale variations, background clutter and fast motion.

Correlation filters have recently attracted a great attention due to their rapid speeds of calculation and robust tracking performance [24, 27, 28, 29]. Bolme et al. [6] proposed an adaptive correlation filer, called MOSSE, for producing ASEF-like filters by fewer training images. Henriques et al. [35] extended the correlation-filter-based trackers to kernel-based training, called CSK method, to utilize a circulant structure of one image patch to conduct dense sampling, and then improved the KCF tracker [19] by using multi-channel inputs and HOG descriptors. Danelljan et al. [14] developed the DSST method handling scale changes of a target, and Choi et al. [10] proposed a spatially attentional weight map to weight various correlation filters. Ma et al. [30] used a correlation filter as a short-term tracker and an online random fern classifier for re-detection as a long-term memory system.

Although achieved the appealing results both in precision and success rate, these correlation-filter-based trackers cannot deal with fast motions and scale variations well. For example, although two correlation-filter-based trackers, namely SCT[10] and KCF [19], have achieved state-of-the-art results and have beaten all other attended trackers in terms of accuracy in the OTB-2013 dataset [43], they fail to track a target object when partial occlusions or illumination changes occur.

To deal with the above issues, we propose a novel and an ensemble tracker, which first builds weak trackers by applying structural correlation filters, and then integrate all weak trackers into one stronger tracker using a reliability evaluation and a multi-task Gaussian particle filter. Each weak tracker is treated as an expert and the weights of all experts are computed via their confidence maps. Through the reliability evaluation, the weak trackers provide weak decisions for the Gaussian particle filter to make a final decision. The tracking result in the current frame is interred by the weights of the Gaussian particle filter. Particles respectively calculate two results using HOG and gray-normalization

features, and therefore the Gaussian particle filter conducts a multi-task tracking for making a strong decision.

The contributions in this paper are summarized as follows.

- A novel ensemble algorithm is proposed to combine structural correlation filters and a Gaussian particle filter into a single stronger tracker.
- A large-scale search scope algorithm using for multiple structural correlation filters is proposed, considering spatial geometric relations between target locations in consecutive frames and therefore making the search reliable.
- Extensive experiments in the OTB-2013 benchmark dataset [43] with 50 challenging sequences and 11 various attributes to demonstrate the outperformance of the proposed method in comparison with 16 state-of-the-art trackers.

2. Related Work

A comprehensive tracking review can be found in [43, 44, 36]. In this section, we discuss the methods closely related to this work, mainly regarding multi-task correlation filters, Gaussian particle filters and Average Peak-to-Correlation Energy (APCE).

2.1. Structural Correlation Filters

Qi et al. [34] described the weak correlation filers on CNN features in each layer and Liu et al. [27] proposed the concept of the structural correlation filter.

In [34], $X^k \in \mathbb{R}^{P\times Q\times D}$ denote the feature map extracted from the k-th convolutional layer with Gaussian function label $Y \in \mathbb{R}^{P\times Q}$. Let $\mathcal{X}^k = \mathcal{F}(X^k)$ and $\mathcal{Y} = \mathcal{F}(Y)$, where $\mathcal{F}(\cdot)$ represents the discrete Fourier transformation (DFT). The objective function of correlation filter method [34] can be extended into its k-th filter modeled as

$$\mathcal{W}^k = \arg\min_{\mathcal{W}} \|\mathcal{Y} - \mathcal{X}^k \cdot \mathcal{W}\|_F^2 + \lambda\|\mathcal{W}\|_F^2, \quad (1)$$

where

$$\mathcal{X}^k \cdot \mathcal{W} = \sum_{d=1}^{D} \mathcal{X}^k_{*,*,d} \bigodot \mathcal{W}_{*,*,d}. \quad (2)$$

Here, the symbol $\bigodot$ is the element-wise product.

The optimization problem in Eq. 1 has a simple closed form solution, which can be efficiently computed in the Fourier domain by

$$\mathcal{W}^k_{*,*,d} = \frac{\mathcal{Y}}{\mathcal{X}^k \cdot \mathcal{X}^k + \lambda} \bigodot \mathcal{X}^k_{*,*,d}. \quad (3)$$

Given the testing data T^k from the output of the k-th layer, we first transform it to the Fourier domain $\mathcal{T}^k = \mathcal{F}(T^k)$, and then the responses can be computed by

$$S^k = \mathcal{F}^{-1}(\mathcal{T}^k \cdot \mathcal{W}^k), \quad (4)$$

where $\mathcal{F}^{-1}$ denotes the inverse of DFT.

The k-th weak tracker outputs the target position with the largest response

$$(x^k, y^k) = \arg\max_{x',y'} S^k(x', y'). \quad (5)$$

2.2. Gaussian Particle Filters (GPF)

Kotecha and Djuric [23] introduced the Gaussian Particle Filter (GPF), which is used for tracking filtering and predictive distributions encountered in Dynamic State-Space models (DSS) [18]. The DSS model represents the time-varying dynamics of an unobserved state variable. GPF is based on the particle filtering and Gaussian filtering concepts. Gaussian filters provide Gaussian approximations to the filtering and predictive distributions, and they include Extended Kalman Filter (EKF) [21] and its variations [20, 31, 37, 42]. Unlike EKF, which assumes that predictive distributions are Gaussian and employs linearization of the functions in the process and observation equations, GPF updates the Gaussian approximations using particles. GPF only propagates the posterior mean and covariance of an unobserved state variable in a DSS model, and essentially importance sampling makes the procedure simple.

Particle filters [16, 32, 33] use sequential importance sampling (SIS) [26] to update the posterior distributions. GPF is quite similar to SIS filters by the fact that importance sampling is used to obtain particles. However, a phenomenon called sample degeneration occurs wherein only a few particles representing the distribution have significant weights. A procedure called resampling [25] has been introduced to mitigate this problem, but it may give limited results and may be computationally expensive. Since GPF approximates posterior distributions as Gaussians, unlike the SIS filters, particle resampling is not required. This results in a reduced complexity of GPF as compared with SIS with resampling and is a major advantage. Furthermore, Berzuini et al. [5] reported that particle filters with resampling also had bias due to resampling, and resampling in SIS filters is a nonparallel operation. Fortunately, resampling would never occur in GPF simulation examples, and the particle filters in GPF are amenable to parallel implementation. Therefore, GPF is more amenable for fully parallel implementation in very large scale integration (VLSI) than SIS.

Simulation results are presented to demonstrate the versatility and improved performance of GPF over conventional Gaussian filters and the lower complexity of GPF than the known particle filters. However, the parallelizibility of GPF and the absence of resampling makes it convenient for VLSI implementation and, hence, feasible for practical real-time applications.

2.3. Ensemble trackers

Multiple component trackers have been combined with hand-crafted features to develop ensemble tracking methods [1, 2, 41] for visual tracking. For example, several ensemble methods [1, 2] using a boosting framework [15] train each component weak tracker to classify foreground objects and background. In [41], Wang and Yeung used a conditional particle filter to infer a target's position and the reliability of each component tracker. Qi et al. [34] treated tracking as a decision-theoretic online learning task and the tracked target was inferred by using the decisions from multiple expert trackers. Similar to [34], we consider visual tracking as a decision-theoretic online learning task [9], and use it in the structure of multiple correlation filters combined with a Gaussian particle filter. That is, in every round, each correlation filter makes a decision and the final decision is determined by a Gaussian particle filter.

2.4. Average Peak-to-Correlation Energy (APCE)

For correlation filters, the peak value F_{max} denotes the maximum response score of a response map. To measure the fluctuation degree of a response map and the reliability degree of a detected object, Wang et al. [40] proposed an average peak-to-correlation energy (APCE) which is defined as

$$APCE = \frac{|F_{max} - F_{min}|^2}{\text{mean}\left(\sum_{w,h}(F_{w,h} - F_{min})^2\right)} \tag{6}$$

where F_{max}, F_{min} and $F_{w,h}$ denote the maximum, minimum and the w-th row and h-th column elements of $F(s, y; w)$. APCE indicates the fluctuated degrees of response maps and the confidence degrees of the tracked results. For sharper peaks and less noise, in the case that the target fully appearing in a tracking region, APCE will become greater and the response map will become smoother except for only one sharp peak. On the other hand, APCE will be small if an object is occluded or missing.

3. Proposed Algorithm

In this section, we present the combination of structural correlation filters with a multi-task Gaussian particle filter for ensemble tracking, namely KCF-GPF. Different from the KCF method [19, 35] that learns a single correlation filter in a fixed-size area, KCF-GPF is proposed to construct multiple weak correlation filters in a more reliable search scope for dealing with fast motion issues and bound effects in the conventional correlation filters. The Gaussian particle filter jointly learns particle weights based on different features to make a stronger tracker by a multi-task method. Furthermore, our tracker can effectively handle scale variations via the sampling strategy of a Gaussian particle filter. Overall, the proposed ensemble method will achieves the following two goals: 1) weak expert trackers are tuned to separate a foreground object from background and 2) the ensemble as a whole ensures the temporal coherence of each part of the tracker.

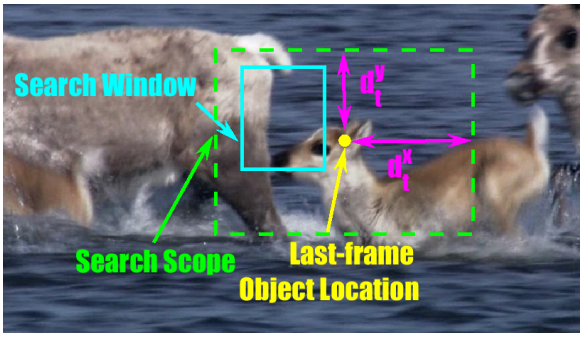

Figure 2. Illustration of the target tracking using the weak structural correlation filter in sequence **Deer** from OTB-2013 dataset [43]. The search scope is based on the max historical moving distance of the object in horizontal and vertical directions at time t, namely d_t^x and d_t^y, and considers the last-frame object location as the center. A search window slides in the search scope to enclose a sample corresponding to a weak correlation filter.

3.1. Weak Structural Correlation Filter

A conventional correlation filter has bound effects [13] during a target tracking and it interferes with the progress of target detection with fast motions. For this, we extend the conventional search window for a single tracker to a large-scale search scope for multiple trackers, and exploit spatial-geometric relations between target locations in consecutive frames to make the search scope reliable.

The feature map X^k in Eq. 1 is extracted from an image patch which is sampled by a search sliding window (see Figure. 2), where d_t donates the maximum historical-moving-distance of a tracked target in Eq. 17 at time t and the object location in the previous frame is the center of the search scope. Besides it, we respectively use d_t^x and d_t^y to represent the horizontal and vertical moving distances at time t.

Given the initial confidence weights of all weak experts, in the current round, a further decision is made based on the expert tracker with the greatest weight. In the visual tracking scenario, it is natural to treat each KCF tracker as an expert and then predict the target position at time t by

$$(x_t^*, y_t^*) = (x_t^k, y_t^k) \cdot [1 - \text{sign}(|w_t^k - \max w_t^k|)], \tag{7}$$

where w_t^k is the weight of expert k and $\sum_{k=1}^{K} w_t^k = 1$, $|\cdot|$ denotes the absolute value and sign represents the signum function.

Kernel Selection: We choose the Gaussian kernel in the existing correlation filter tracker [19].

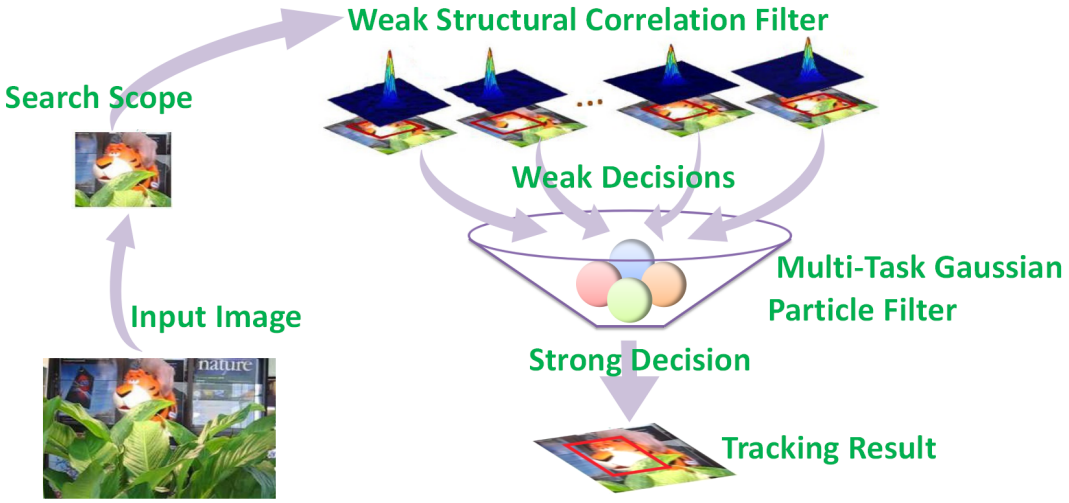

Figure 1. Flowchart of the proposed algorithm. The proposed algorithm consists of three components: 1) constructing multiple weak trackers using correlation filters where each one is trained using HOG features and makes a weak decision (Section 3.1); 2) evaluating the reliability degree of each weak decision via maximum response score and APCE measure and the most reliable decision is used for the next step (Section 3.2); 3) constructing a stronger tracker and making a final decision via a multi-task Gaussian particle filter which takes the most reliable decision into account (Section 3.3).

Feature Representation: Similar to [19], we use HOG features with 31 bins. However, our tracker is quite generic and any dense feature representation with arbitrary dimensions can be incorporated.

Compared to the HDT [34] and SCF [27] methods which are similar to the proposed weak structural correlation filter, we demonstrate differences among these approaches as follows.

1. HDT uses CNN features, while SCF and KCF-GPF are based on HOG features.

2. The features of HDT are extracted from one layer to build a weak tracker, and the part-based correlation filter SCF samples several parts of a target object to construct features, while KCF-GPF samples via sliding window in a search scope which is subject to the maximum historic-moving-displacement of the tracked target over time t.

3. In HDT, the target position is made based on the weighted decisions of all experts and SCF solves the optimization problem using the fast first-order Alternating Direction Method of Multipliers (ADMM) [7], while KCF-GPF exploits Eq. 7 to infer the ultimate target position.

3.2. Reliability Degree of Correlation Filter

The peak value and the fluctuation of the response map can reveal the confidence degree about the tracking results to some extent. The ideal response map should have only one sharp peak and be smooth in all other areas when the detected target is extremely matched to the correct target. The sharper the correlation peaks are, the better the location accuracy is. Otherwise, the whole response map will fluctuate intensely, and its pattern is significantly different from normal response maps. If we continue to use uncertain samples to update the tracking model, it would be corrupted mostly.

Inspired by [40], we evaluate the stability of the k-th expert with two criteria. The first criterion is the maximum response score F_{max} of the response map defined as

$$F_{max} = \max_{x',y'} S^k(x', y') \tag{8}$$

where $S^k(x', y')$ is referred to Eq. 4 and Eq. 5.

The second criterion is called average peak-to-correlation energy (APCE) measure which is defined in Eq. 6.

3.3. Multi-task Gaussian Particle Filter

The proposed multi-task Gaussian particle filter at time t approximates the posterior mean $\boldsymbol{\mu}_t$ and covariance $\boldsymbol{\Sigma}_t$ of the unknown state variable $\mathbf{x}_t$ using Bayesian importance sampling.

We draw samples from the importance function $\pi(\cdot)$ at time t using

$$\pi(\mathbf{x}_t|\mathbf{y}_{0:t}) = \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), \tag{9}$$

and denote them as $\{\mathbf{x}_t^j\}_{j=1}^M$. Here, $\mathbf{y}_{0:t}$ is the observations over time t, and $\mathcal{N}(\cdot)$ represents a Gaussian function. Note that, in Eq. 9, $\boldsymbol{\mu}_t$ is equal to Eq. 7 and $\boldsymbol{\Sigma}_1$ is chosen based on prior information.

The respective weights are computed by

$$w_t^j = \frac{p(\mathbf{y}_t|\mathbf{x}_t^j)\mathcal{N}(\mathbf{x}_t = \mathbf{x}_t^j; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)}{\pi(\mathbf{x}_t^j|\mathbf{y}_{0:t})}, \tag{10}$$

where the distribution $p(\mathbf{y}_t|\mathbf{x}_t^j)$ represents the observation equation $\mathbf{y}_t$ conditioned on the unknown state variable $\mathbf{x}_t^j$ at time t.

Eq. 10 can be rewritten as follows from Eq. 9:

$$w_t^j \propto p(\mathbf{y}_t|\mathbf{x}_t^j). \tag{11}$$

Then, we set $p(\mathbf{y}_t|\mathbf{x}_t^j) = |f_t^* - f(\mathbf{x}_t^j)|$, where $|\cdot|$ denotes the absolute value and f_t is the function of features, where f_t^* represents the features of the template at time t. Hence, each Gaussian particle weight can be calculated with

$$w_t^j \propto |f_t^* - f(\mathbf{x}_t^j)| \tag{12}$$

Normalize the weights as

$$\tilde{w}_t^j = \frac{w_t^j}{\sum_{j=1}^M w_t^j}. \tag{13}$$

In this paper, we adopt HOG for tackling deformation variations and gray normalization of raw pixel values from sample images for handling illumination changes, and the j-th features are represented by $f_{t(hog)}^j$ and $f_{t(norm)}^j$ at time t, respectively . Hence, we get the corresponding weights $\tilde{w}_{t(hog)}^j$ and $\tilde{w}_{t(norm)}^j$.

For the multi-task Gaussian particle filter, we jointly define the similarities of respective samples as

$$\bar{w}_t^j = \theta \cdot \tilde{w}_{t(hog)}^j + (1-\theta) \cdot \tilde{w}_{t(norm)}^j, \tag{14}$$

where learning rate parameter θ is a constant value in this paper.

The mean and covariance are estimated by

$$\boldsymbol{\mu}_t = \sum_{j=1}^M \bar{w}_t^j \mathbf{x}_t^j, \tag{15}$$

$$\boldsymbol{\Sigma}_t = \sum_{j=1}^M \bar{w}_t^j (\boldsymbol{\mu}_t - \mathbf{x}_t^j)(\boldsymbol{\mu}_t - \mathbf{x}_t^j)^H, \tag{16}$$

where H represents the Hermitian Matrix.

The maximun historical-moving-distance of the tracked target at time t is defined as

$$d_t = \max_{i=1}^{t}(|\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i-1}|). \tag{17}$$

Here, we set $d_1 = 0$, and $|\cdot|$ denotes the function for absolution values;

3.4. KCF-GPF Tracker

Figure 1 illustrates the flowchart of the proposed algorithm. Based on the structural correlation filter and a Gaussian particle filter, we propose a KCF-GPF tracker. The first step generates K weak correlation filters. The second step is to make weak decisions via the weak structural correlation filter.s The third step is to evaluate the reliability degrees of weak decisions. Finally, the optimal decision is made using a multi-task Gaussian particle filter.

To update the KCF-GPF for visual tracking, we adopt an incremental strategy in the current frame to update the template in Eq. 12 by

$$f_t^* = \rho \cdot f_{t-1}^* + (1-\rho) \cdot f_{t-1}(\boldsymbol{\mu}_{t-1}), \tag{18}$$

where learning rate parameter ρ is a constant value in this paper.

An overview of the proposed method is summarized in Algorithm 1.

Algorithm 1: KCF-GPF tracking algorithm

Input: Frames $\{\mathbf{I}_t\}_1^T$;
Output: Target location of each frame $\boldsymbol{\mu}_t$.

```
for Time t = 1 : T do
    if t = 1 then
        Initialize the target location (x_1^*, y_1^*);
        Crop interested image region;
        Initiate K weak correlation filters using Eq. 3;
    else
        Compute each correlation filter's response
        using Eq. 4;
        Find target position predicted by each weak
        tracker using Eq. 5;
        Evaluate the reliability degrees APCE and
        F_max via Eq. 6 and Eq. 8;
        if max_{k=1}^{K} APCE and max_{k=1}^{K} F_max
        satisfy the condition then
            Draw samples from Eq. 9;
        else
            Draw samples from Eq. 9 where Σ_t = 3;
        end
        Compute the ultimate position using Eq. 15;
        Update the template via Eq. 18;
    end
end
```

4. Experiments

4.1. Experimental Setups

Implementation Details. The conventional features used for KCF-GPF are composed of HOG features and gray

Table 1. Parameters of KCF-GPF

Part of Track	parameters	values
KCF	padding	1.2
	Feature bandwidth σ	0.5
	Adaptation rate	0.045
GPF	Sample numbers N	200
	Learning rate θ (Eq. 14)	0.65
	Learning rate ρ (Eq. 18)	0.8

normalization features. Our tracker is implemented on MATLAB in a PC with a 2.80 GHz CPU and runs faster than 21 FPS in Table 2. Our tracker requires few parameter settings, reported in Table 1, where 'padding' (referring to KCF [19]) means the magnification of the image region samples relative to the target bounding box.

Datasets. Our method is evaluated in the OTB-2013 dataset [43] consisting of 50 sequences. The images are annotated with ground truth bounding boxes and 11 various visual attributes include scale variation, out of view, out-of-plane rotation, low resolution, in-plane rotation, illumination, motion blur, background clutter, occlusion, deformation, and fast motion.

Evaluation Metrics. We compare the proposed method with the 16 state-of-the-art tracking methods using evaluation metrics and code provided by the respective benchmark dataset. For testing on OTB-2013, we employ the one-pass evaluation (OPE) and use two metrics: precision and success plots. The precision metric computes the rate of frames whose center location is within some certain distance from the ground truth location. The success metric computes the overlap ratio between the tracked and ground truth bounding boxes. In the legend, we report the area under curve (AUC) of success plot and precision score at a 20 pixel threshold (PS) corresponding to the one-pass evaluation for each tracking method.

4.2. Comparison with State-of-the-Art

We evaluate KCF-GPF in the OTB-2013 dataset [43] and compare it with 16 state-of-the-art trackers including LMCF [40], CFNet [39], CFN [11], CFN_ [11], CNT [46], BIT [8], SINT [38], SCT [10], Staple [3], SiamFC [4], SRDCF [13], DSST [12], MEEM [45], KCF [19], TLD [22] and Struck [17]. Among them, LMCF, CFN, CFN_, Staple, SRDCF, KCF, DSST, CFNet and SCT are CF based algorithms; SiamFC, SINT, CFNet, CNT and BIT are convolutional networks based algorithms; MEEM is developed based on regression and multiple trackers; TLD is based on an ensemble classifier; and Struck is structured SVM based methods.

The characteristics and tracking results are summarized in Table 2. The mean FPS here is estimated on all sequences in the OTB-2013 and achieves 21.3 fps satisfying the requirement of real-time capability. LMCF achieves the second best performance in terms of the success metric and SINT shows the second best performance in terms of precision metric. Figure 5 illustrates the precision and success plots of all trackers under all challenging attributes in the OTB-2013. KCF-GPF is also superior to other up-to-date trackers with precision and success evaluation metrics in the OTB-2013 benchmark.

For detailed analyses, we also evaluate KCF-GPF with these state-of-the-art trackers on various challenging attributes in the OTB-2013 benchmark dataset and the results are shown in Figure 3 and Figure 4. The results demonstrate that KCF-GPF is ranked on top three in each attribute and achieves the best performances in the general success plots. Besides that, the proposed method outperforms other trackers in terms of deformation and out-of-plane rotation attributes.

4.3. Qualitative Comparison

To demonstrate the effect of the proposed KCF-GPF tracking algorithm, we make a qualitative comparison with above state-of-the-art trackers in the OTB-2013 benchmark dataset with 11 different attributes. As shown in Figure 6, all trackers perform well overall, but the existing trackers have the following drawbacks. The SCT does not perform well under scale variations (Liquor, Woman, and Dog1). The CFNet cannot handle occlusion (Lemming, Skating1, Subway, Singer2, Suv, Liquor, Woman and Soccer), deformation(Skating1, Subway, Singer2, Suv and Woman) ,out-of-plane rotation (Lemming, Skating1, Singer2, Liquor, Woman and Soccer) and background clutters (Skating1, Subway, Singer2, Suv, Liquor and Soccer). The KCF drifts when there are illumination variations (Shaking, Lemming and Woman), scale variations (Shaking, Lemming, Woman and Dog1), out-of-plane rotations (Shaking, Lemming, Woman and Soccer), and fast motions (Woman and Soccer). The TLD and Struck methods drift when target objects undergo illumination changes (Shaking, Skating1, Singer2 and Soccer), heavy occlusion (Lemming, Subway, Singer2, Suv, Liquor, Woman and Soccer) and scale variations (Lemming and Dog1). Overall, the proposed KCF-GPF tracker performs the best against the existing trackers in tracking objects on these challenging sequences.

5. Conclusion

In this paper, we have proposed a novel tracker combining multiple structural correlation filters with a multi-task gaussian particle filter, namely KCF-GPF, to construct a strong tracker for ensemble tracking. The proposed method takes multiple correlation filters as weak expert trackers, and exploits spatial-geometric relations between target locations in consecutive frames to provide weak decisions in

Table 2. Tracking results of all 17 evaluated trackers over all 50 sequences using OPE evaluation in the OTB-2013. The entries in red denote the best results and the ones in blue indicate the second best.

Trackers / OPE	LMCF[40]	CFNet[39]	CFN[11]	CFN_ [11]	CNT[46]	BIT[8]	SINT[38]	SCT[10]	Staple[3]
precision	0.842	0.803	0.813	0.784	0.723	0.816	0.851	0.836	0.793
success	0.800	0.775	0.675	0.630	0.656	0.745	0.791	0.730	0.754
Trackers / OPE	**SiamFC[4]**	**SRDCF[13]**	**DSST[12]**	**MEEM[45]**	**KCF[19]**	**TLD[22]**	**Struck[17]**	**KCF-GPF(ours)**	**mean FPS(ours)**
precision	0.809	0.838	0.740	0.840	0.740	0.608	0.656	0.857	21.3
success	0.783	0.781	0.670	0.706	0.623	0.521	0.559	0.805	

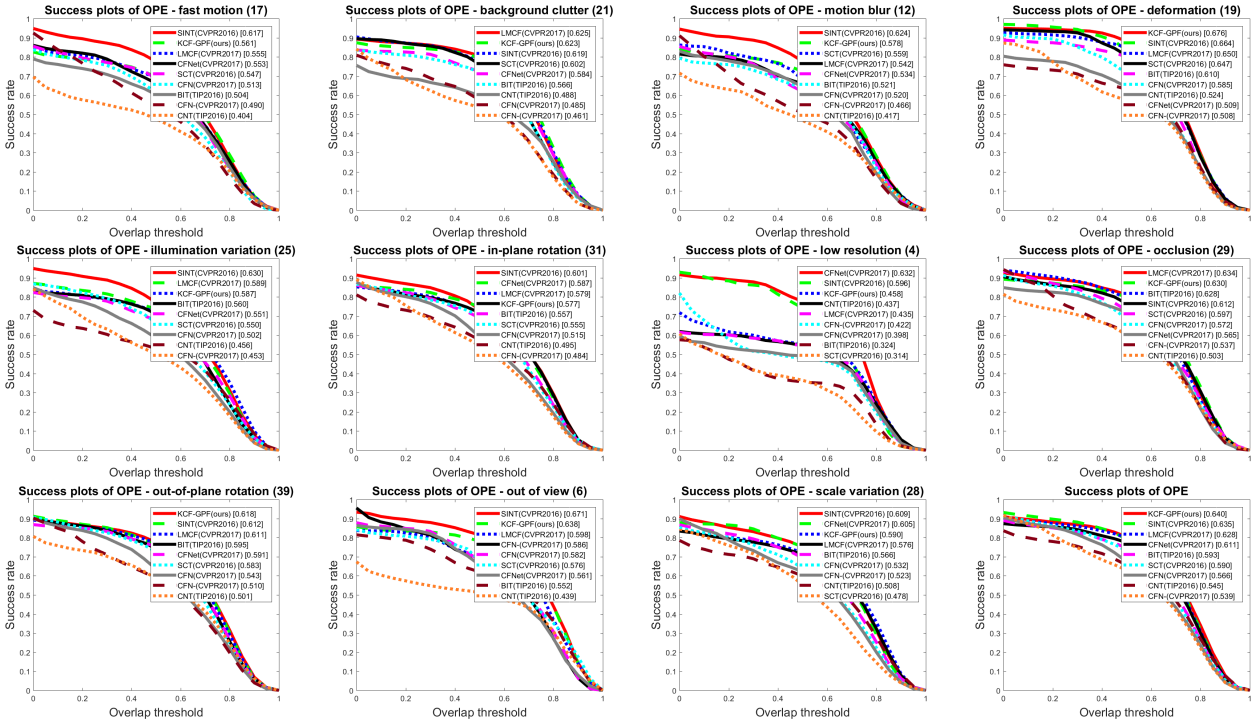

Figure 3. Success plots over all 50 sequences using OPE evaluation in the OTB-2013 dataset. The evaluated trackers are LMCF, CFNet, CFN, CFN_, CNT, BIT, SINT, SCT and KCF-GPF. All 11 tracking challenges include scale variation, out of view, out-of-plane rotation, low resolution, in-plane rotation, illumination, motion blur, background clutter, occlusion, deformation, and fast motion. The numbers in the legend indicate the average AUC scores for success plots. Our KCF-GPF method performs favorably against the state-of-the-art trackers.

a reliable search scope. The reliability degrees of weak decisions are introduced in experiments for the GPF to make a strong decision. As a result, it not only has the advantages of the existing correlation filter trackers, such as, computational efficiency and robustness, but also can deal with scale variations by the sampling strategy of a GPF. The proposed KCF-GPF tracking algorithm outperforms 16 state-of-the-art methods over all 50 sequences in the OTB-2013 benchmark in terms of qualitative and quantitative evaluations.

References

[1] S. Avidan. Ensemble tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):261–271, 2007. 3

[2] Q. Bai, Z. Wu, S. Sclaroff, M. Betke, and C. Monnier. Randomized ensemble tracking. In *IEEE International Conference on Computer Vision*, pages 2040–2047, 2013. 3

[3] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr. Staple: Complementary learners for real-time tracking. 38(2):1401–1409, 2015. 6, 7

[4] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr. Fully-convolutional siamese networks for object tracking. pages 850–865, 2016. 6, 7

[5] C. Berzuini, N. G. Best, W. R. Gilks, and C. Larizza. Dynamic conditional independence models and markov chain monte carlo methods. *Journal of the American Statistical Association*, 92(440):1403–1412, 1997. 2

[6] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui. Visual object tracking using adaptive correlation filters.

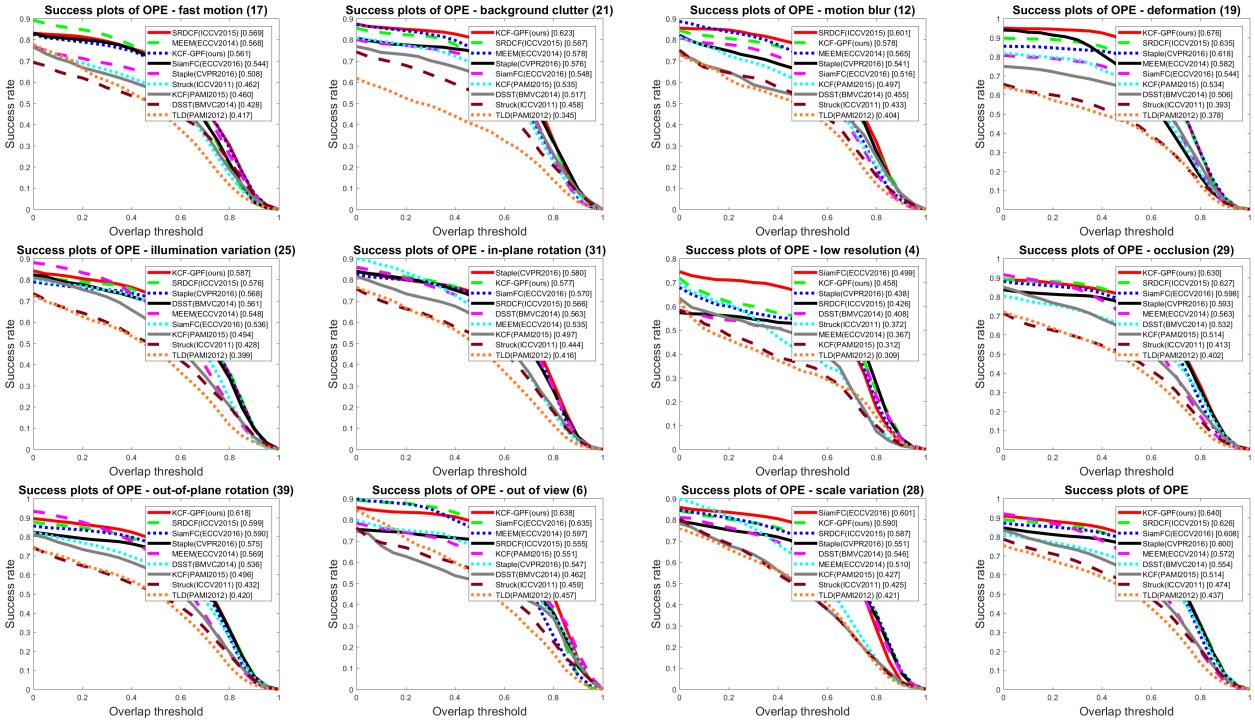

Figure 4. Success plots over all 50 sequences using OPE evaluation in the OTB-2013 dataset. The evaluated trackers are Staple, SiamFC, SRDCF, DSST, MEEM, KCF, TLD, Struck and KCF-GPF. All 11 tracking challenges include scale variation, out of view, out-of-plane rotation, low resolution, in-plane rotation, illumination, motion blur, background clutter, occlusion, deformation, and fast motion. The numbers in the legend indicate the average AUC scores for success plots. Our KCF-GPF method performs favorably against the state-of-the-art trackers.

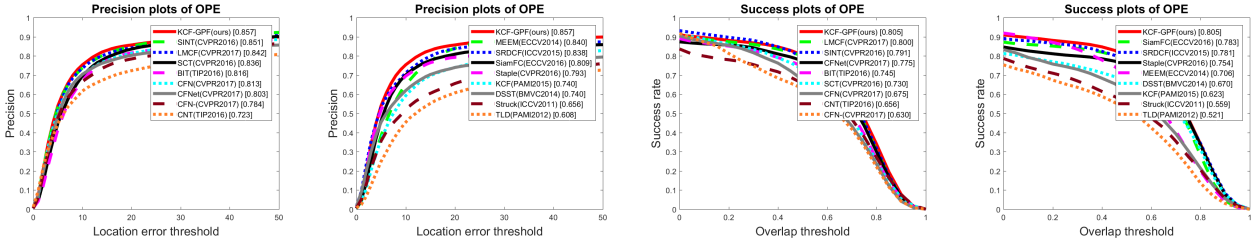

Figure 5. Precision and success plots over all 50 sequences using OPE evaluation in the OTB-2013 dataset. The numbers in the legend indicate the average precision scores for precision plots and the average AUC scores for success plots. Our KCF-GPF method performs favorably against the state-of-the-art trackers.

In *Computer Vision and Pattern Recognition*, pages 2544–2550, 2010. 1

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. 4

[8] B. Cai, X. Xu, X. Xing, K. Jia, J. Miao, and D. Tao. Bit: Biologically inspired tracker. *IEEE Transactions on Image Processing*, 25(3):1327–1339, 2016. 6, 7

[9] K. Chaudhuri, Y. Freund, and D. Hsu. A parameter-free hedging algorithm. *Computer Science*, pages 297–305, 2009. 3

[10] J. Choi, H. J. Chang, J. Jeong, Y. Demiris, and Y. C. Jin. Visual tracking using attention-modulated disintegration and integration. In *Computer Vision and Pattern Recognition*, pages 4321–4330, 2016. 1, 6, 7, 9

[11] J. Choi, H. J. Chang, S. Yun, T. Fischer, Y. Demiris, and Y. C. Jin. Attentional correlation filter network for adaptive visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 6, 7

[12] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg. Accurate scale estimation for robust visual tracking. In *British Machine Vision Conference*, pages 65.1–65.11, 2014. 6, 7

[13] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg. Learn-

Figure 6. Comparisons of the proposed tracker with the state-of-the-art trackers (SCT [10], CFNet [39], KCF [19], [22] and Struck [17]) in our evaluation on 10 challenging sequences (from left to right and top to down are **Shaking**, **Lemming**, **Skating1**, **Subway**, **Singer2**, **Suv**, **Liquor**, **Woman**, **Soccer**, **Dog1**, respectively).

ing spatially regularized correlation filters for visual tracking. In *IEEE International Conference on Computer Vision*, pages 4310–4318, 2015. 3, 6, 7

[14] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg. Discriminative scale space tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(8):1561–1575, 2016. 1

[15] Y. Freund and R. E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory*, pages 23–37, 1995. 3

[16] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P. J. Nordlund. Particle filters for positioning, navigation, and tracking. *IEEE Transactions on Signal Processing*, 50(2):425–437, 2002. 2

[17] S. Hare, A. Saffari, and P. H. S. Torr. Struck: Structured output tracking with kernels. In *IEEE International Conference on Computer Vision*, pages 263–270, 2012. 6, 7, 9

[18] P. J. Harrison and C. F. Stevens. Bayesian forecasting. discussion. *Journal of the Royal Statistical Society*, 38, 1976. 2

[19] J. F. Henriques, C. Rui, P. Martins, and J. Batista. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):583–596, 2015. 1, 3, 4, 6, 7, 9

[20] Jazwinski and AndrewH. *Stochastic processes and filtering theory*. Academic Press,, 1970. 2

[21] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004. 2

[22] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1409, 2012. 6, 7, 9

[23] J. H. Kotecha and P. M. Djuric. Gaussian particle filtering. *IEEE Transactions on Signal Processing*, 51(10):2592–2601, 2003. 2

[24] Y. Li, J. Zhu, and S. C. H. Hoi. Reliable patch trackers: Robust visual tracking by exploiting reliable patches. In *Computer Vision and Pattern Recognition*, pages 353–361, 2015. 1

[25] J. S. Liu and R. Chen. Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044, 1998. 2

[26] J. S. Liu, R. Chen, and T. Logvinenko. *A Theoretical Framework for Sequential Importance Sampling with Resampling*. Springer New York, 2001. 2

[27] S. Liu, T. Zhang, X. Cao, and C. Xu. Structural correlation filter for robust visual tracking. In *Computer Vision and Pattern Recognition*, pages 4312–4320, 2016. 1, 2, 4

[28] T. Liu, G. Wang, and Q. Yang. Real-time part-based visual tracking via adaptive correlation filters. In *Computer Vision and Pattern Recognition*, pages 4902–4912, 2015. 1

[29] C. Ma, J. B. Huang, X. Yang, and M. H. Yang. Hierarchical convolutional features for visual tracking. In *IEEE International Conference on Computer Vision*, pages 3074–3082, 2015. 1

[30] C. Ma, X. Yang, C. Zhang, and M. H. Yang. Long-term correlation tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5388–5396, 2015. 1

[31] J. Mendel. Optimal filtering. *IEEE Transactions on Automatic Control*, 25(3):615–616, 1980. 2

[32] R. V. D. Merwe, A. Doucet, N. D. Freitas, and E. Wan. The unscented particle filter. *Advances in Neural Information Processing Systems*, 13:584–590, 2001. 2

[33] K. Nummiaro, E. Koller-Meier, and L. V. Gool. An adaptive color-based particle filter. *Image and Vision Computing*, 21(1):99–110, 2003. 2

[34] Y. Qi, S. Zhang, L. Qin, H. Yao, Q. Huang, J. Lim, and M. H. Yang. Hedged deep tracking. In *Computer Vision and Pattern Recognition*, pages 4303–4311, 2016. 2, 3, 4

[35] C. Rui, P. Martins, and J. Batista. Exploiting the circulant structure of tracking-by-detection with kernels. In *European Conference on Computer Vision*, pages 702–715, 2012. 1, 3

[36] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah. Visual tracking: An experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1442–1468, 2014. 2

[37] H. W. Sorenson. Recursive estimation for nonlinear dynamic systems. *Bayesian Analysis of Time*, 1988. 2

[38] R. Tao, E. Gavves, and A. W. M. Smeulders. Siamese instance search for tracking. In *Computer Vision and Pattern Recognition*, pages 1420–1429, 2016. 6, 7

[39] J. Valmadre, L. Bertinetto, J. F. Henriques, A. Vedaldi, and P. H. S. Torr. End-to-end representation learning for correlation filter based tracking. 2017. 1, 6, 7, 9

[40] M. Wang, Y. Liu, and Z. Huang. Large margin object tracking with circulant feature maps. 2017. 1, 3, 4, 6, 7

[41] N. Wang and D. Y. Yeung. Ensemble-based tracking: aggregating crowdsourced structured time series data. In *International Conference on International Conference on Machine Learning*, pages II–1107, 2014. 3

[42] G. Welch and G. Bishop. An introduction to the kalman filter. *University of North Carolina at Chapel Hill*, 8(7):127–132, 2010. 2

[43] Y. Wu, J. Lim, and M. H. Yang. Online object tracking: A benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2411–2418, 2013. 1, 2, 3, 6

[44] A. Yilmaz. Object tracking: A survey. *Acm Computing Surveys*, 38(4):13, 2006. 2

[45] J. Zhang, S. Ma, and S. Sclaroff. Meem: Robust tracking via multiple experts using entropy minimization. In *European Conference on Computer Vision*, pages 188–203, 2014. 6, 7

[46] K. Zhang, Q. Liu, Y. Wu, and M. H. Yang. Robust visual tracking via convolutional networks without training. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 25(4):1779, 2016. 1, 6, 7